Psychological Crisis Prediction in College Students Using a Stacked Ensemble of Random Forest, Logistic Regression, and AdaBoost

Shuying Hao
School of Food and Biology, Jingchu University of Technology, Jingmen, 448000, China,
E-mail: haomm20192@outlook.com

Keywords: psychological crisis, logistic regression, random forest, early warning system, iterative algorithm

Received: September 15, 2025

Recently, campus malignant incidents caused by psychological problems among college students have occurred frequently, causing particularly serious impacts on students, families, and schools. At present, the main method for conducting psychological surveys in major universities is through questionnaire surveys. However, this method can make students with mental illness choose answers that are beneficial to them, resulting in misjudgments by psychologists. In response to this issue, this study proposes a stacked fusion model based on Random Forest (RF), Logistic Regression (LR), and AdaBoost, aiming to improve the accuracy and robustness of psychological crisis detection. A dataset consisting of 200 college students' psychological assessment results, behavior logs, and demographic features (totaling 21 indicators) is collected through game simulation and questionnaire surveys. The model collects student information through game simulation, establishes students' psychological files, and analyzes the files to determine whether students have a psychological crisis. The experiment demonstrated that the average absolute error values of LR + iterative algorithm, RF + iterative algorithm, RF + LR, and mixed models were 0.70, 0.69, 0.68, and 0.67, respectively. The root mean square error values were 0.889, 0.885, 0.881, and 0.879, respectively. When the training set was small, the performance of the hybrid model could reach a higher level, and when the validation set was too large, its performance did not significantly decrease. Among all models, the proposed RLA-Stacking model achieved the highest F1-score of 81.8%, outperforming single models and other fusion variants. The research demonstrates that the proposed hybrid model possesses good performance in forecasting psychological crises among college students.

Povzetek: Predlagan je zloženi fuzijski model (RF + LR + AdaBoost) za zaznavanje psiholoških kriz študentov, ki na vedenjskih in anketnih podatkih doseže najboljši rezultat (F1 = 81,8 %) ter večjo robustnost od posameznih modelov.

1 Introduction

With the national education reform and the expansion of enrollment in higher education institutions, the number of college students has surged. Although college students have more knowledge reserves compared to other members of society, their Mental Health (MH) status is worrying and has become an important issue that cannot be ignored among college students. Due to the fact that college students are generally studying in other places, far away from their families, and lack methods to relieve psychological stress, the accumulated pressure cannot be released, resulting in some college students having long-term psychological problems [1]. The traditional MH survey of college students is conducted through questionnaire surveys. Students with psychological problems may conceal their own situation to change the results of the questionnaire, which leads to many people who originally needed psychological intervention not receiving effective treatment [2]. In response to this issue, this study proposes a Stcaking fusion model based on Random Forest (RF), Logistic Regression (LR), and Adaboost. The model establishes archives based on students' life practices, and uses a college student psychological crisis game simulation warning system that integrates the Stacking model to assess students' psychological status [3-4]. However, the privacy protection of the system is also an important consideration. It is necessary to ensure that the personal information and privacy of college students are fully protected and authorized. The research content is separated into four sections. The first introduces existing research on college students' psychological problems. The second section reviews the main method used in this study. The third section is the model results obtained from the research method, and the results are analyzed. The fourth part summarizes the above research and provides prospects for future research.

Traditional psychological crisis prediction models primarily depend on self-reported questionnaire data or static demographic information, which are limited by subjective bias and response reliability. In contrast, recent machine learning approaches have attempted to infer MH status using social media content or sensor-based activity patterns. However, these methods remain passive in nature and often lack interpretability and behavioral context. To address these limitations, the present research introduces an interactive game simulation-based behavioral data collection mechanism that captures participants' decision-making speed, hesitation, and emotional fluctuations under controlled scenarios. This method can observe implicit psychological signals outside of self-report scales and improve ecological validity. Moreover, the study designs an RLA-Stacking framework, which integrates linear and nonlinear feature representations with iterative optimization to improve robustness and predictive precision.

2 Related work

Recently, psychological problems have become the main sub health problem. Li found that with the advancement of urban-rural integration, migrant children not only face social pressure, but also life pressure. This inevitably affects the physical and MH of children. In response to this issue, they proposed a model based on cloud computing and data mining algorithms to analyze the MH of urban migrant children. The experiment demonstrates that the model can make relatively accurate judgments on children's psychological status [5-6]. Marques et al. found that the outbreak of COVID-19 in 2019 had an essential influence on the MH. They treated college students with mental illness through mobile psychological care programs; The experimental results indicate that mobile psychological care programs have a good effect on alleviating the psychological problems of college students [7]. Sun et al. found that the incidence of MH issues in college students markedly grew recently. Over time, the overall MH status of college students has been deteriorating. They analyzed the reasons for this situation and guided public health prevention work [8]. Chu et al. believed that MH is the basic guarantee for college students to become pillars of talent, and many contemporary college students have psychological problems. Therefore, a detailed study was conducted on the MH status of college students by establishing a MH assessment model. The research results indicate that the proposed model can markedly assess the MH status of college students through psychological data [9].

MH status of college students needs to be addressed. Liang et al. proposed a computational intelligence-based MH quality evaluation method. This method can estimate the MH quality of college students through their behavior and actions. The experiment demonstrates that the model possesses excellent reliability. The estimation outcomes can markedly reflect the psychological issues such as stress, and anxiety of college students [10]. Jianhuan et al. believed that the MH status of college students is an important component of higher education. To better understand the MH status of college students, a prediction system for MH based on BP neural network was proposed. The experiment demonstrates that the error between the predicted and measured values of the model is small. It can provide support for MH education for college students [11]. Shao et al. presented a machine learning algorithm for the RF algorithm to address the problem of not being able to quickly obtain crop coefficient values in estimating crop evapotranspiration in the field. This method can quickly and accurately obtain field crop coefficients. Experimental results show that the combination of unmanned aerial vehicle multi-spectral remote sensing technology and RF algorithm exhibits high performance in water allocation and precise irrigation [12]. Zhang et al. presented a permeability model based on BP neural network and RF to explore the permeability change with pore structure parameters of carbonate rocks. The experiment demonstrates that the model possesses high performance [13]. Liu et al. believed that a reasonable risk identification tool will help solve investors' irrational problems. In response to this problem, they proposed a risk identification tool based on the RF algorithm. The experiment demonstrates that the risk identification system can provide effective identification performance to investors [14].

In summary, many scholars have conducted research on the MH of college students and have achieved certain results; However, most scholars overlook the students' wishes when conducting surveys and research, which can make them unwilling to reveal their psychological problems; Therefore, this study establishes archives based on students' life practices, and uses a college student psychological crisis game simulation warning system that integrates the Stacking model to assess students' psychological status. This avoids students intentionally concealing their MH status in the questionnaire survey, while also providing some protection for their privacy.

Table 1: Related works

Research	Method	Research Content	Dataset Description	Result Metric	Reference
Li (2021)	Cloud Computing + Data Mining	Analyze migrant children's psychological state	Urban migration dataset	Accuracy ≈ 81%	[3]
Marques et al. (2021)	Mobile MH app	COVID impact + intervention effect	University students in Spain	Improved post-intervention MH score	[4]
Chu & Yin (2021)	Clustering Algorithm	MH classification for college students	Survey-based MH scores	Not Reported	[7]
Liang et al. (2022)	Computational Intelligence	Evaluate college students' behavior features	Behavioral datasets	Accuracy ≈ 84%	[8]
Jianhuan et al. (2022)	BP Neural Network	Predict psychological condition	Questionnaire + BP model	MAE ≈ 0.75	[9]

3 Research on psychological crisis game simulation and warning system based on fusion stacking model

The first section of this chapter analyzes the data mining methods used in this study, such as Decision Trees (DTs), RF, iterative algorithms, LR, and psychological crises. The second section establishes a hybrid model through RF, LF, and iterative algorithms. Then, the MH of students is predicted through a mixed model.

3.1 Data mining technology and psychological crisis analysis

Data mining is a process of analyzing a large amount of data, discovering patterns, associations, and trends within it, and extracting valuable information and knowledge. Data mining discovers hidden patterns, trends, and associations from a large amount of data. The goal of data mining is to provide support for decision-making and prediction through data analysis and modeling, such as predictive modeling, association analysis, clustering analysis, and anomaly detection.

Predictive modeling is a technique that uses historical data to construct models to predict future events or outcomes. Predictive modeling is to find suitable mathematical models to predict unknown results by analyzing and understanding patterns and correlations in data. Association analysis is a technique used to discover the correlation between items in a dataset. It analyzes a large amount of data, identifies frequently occurring items or events, and generates association rules based on their correlation. Association rules are rules that describe the association relationship between items. Cluster Analysis (CA) refers to the process of grouping similar data objects into clusters. The goal of clustering analysis is to naturally divide data objects into different groups or clusters based on their similarity without the need for pre-defined categories. CA can help discover potential patterns, structures, and relationships in a dataset. In CA, the similarity of data objects is measured by calculating their distance or similarity. The clustering algorithm determines the similarity between data objects based on the calculation results of distance or similarity, and assigns similar objects to the same cluster. Anomaly point detection is used to identify data points in a dataset that are significantly different from other data points.

DT classification is a method used to solve classification problems. This method employs the idea of recursive segmentation of the training dataset, dividing the dataset into multiple small, pure subsets. Each partition maximizes the classification accuracy by selecting the best attributes. Its structure is shown in Figure 1.

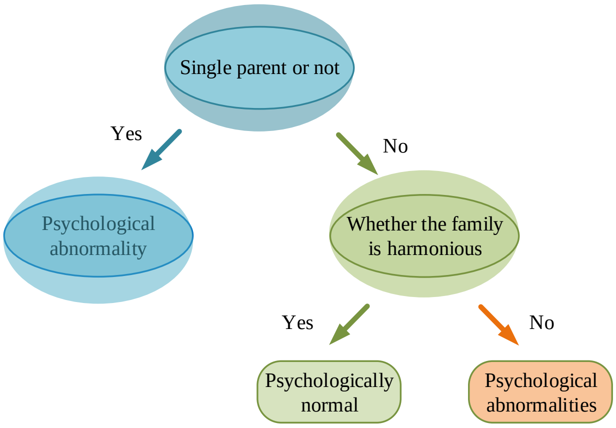


Figure 1: DT structure

Figure 1 shows that the construction process of the DT classification method is similar to the growth process of a tree, with each internal node serving as an attribute test, each branch serving as a test result, and each leaf node serving as a category label [15]. By constructing a DT, the category to which an instance belongs can be predicted based on the input feature values. The DT classification method has strong interpretability, easy understanding, and implementation. The DT classification method is divided into three steps. Feature selection is a standard process of selecting suitable points from numerous features to split. Common measurement methods include information gain, information gain rate, and Gini coefficient. The information entropy is shown in Equation (1).

$$Entropy(t) = -\sum_{i=1}^{c} p(i|t)\log_2 p(i|t) \qquad (1)$$

In equation (1), t represents a node. i represents the category. c represents the number of different categories. $p(i|t)$ serves as the proportion of records belonging to class i in a given node t. The information gain is shown in equation (2).

$$\Delta info = I(parent) - \sum_{j=1}^{k} \frac{N(Vj)}{N} I(Vj) \qquad (2)$$

In equation (2), $I(\square)$ represents the impure degree of a given node. N represents the total number of records on the parent node. k serves as the number of attribute values. $N(Vj)$ represents the number of disciplines associated with the child node Vj. The information gain rate is shown in equation (3).

$$Gain\ ratio = \frac{\Delta info}{SplitInfo} \quad Gain\ ratio = \frac{\Delta info}{SplitInfo} \qquad (3)$$

In equation (3), $\Delta info$ represents information gain. $SplitInfo$ represents partition information, as shown in equation (4).

$$SplitInfo = -\sum_{i=1}^{k} p(v_i)\log_2 p(v_i) \qquad (4)$$

In equation (4), k represents the total number of partitions. $p(v_i)$ represents the proportion of samples occupied by the v_i-th attribute value of the partition attribute in the current node. In addition, the Gini coefficient is shown in equation (5).

$$Gini(t) = 1 - \sum_{i=1}^{c} P[i/t]^2 \qquad (5)$$

The DT generally adopts a greedy strategy, which generates nodes sequentially from top to bottom based on the selected feature criteria until all record attribute values in the same class are the same [16]. The DT classification method is prone to overfitting problems, and in the process of constructing a DT, each subset may contain the same dataset. Therefore, pruning operations are required, which can be divided into pruning first and pruning later. Both pruning methods can effectively solve the overfitting in DT.

RF uses the ensemble learning, which can solve the problem of overfitting in a single DT. By using learning algorithms to generate a single learner on training data, the learning performance of this learner is generally poor. This study uses multiple learners to complement each other's strengths and weaknesses, and cooperates with multiple learners to enhance their performance. The RF model with multiple learners is shown in Figure 2.

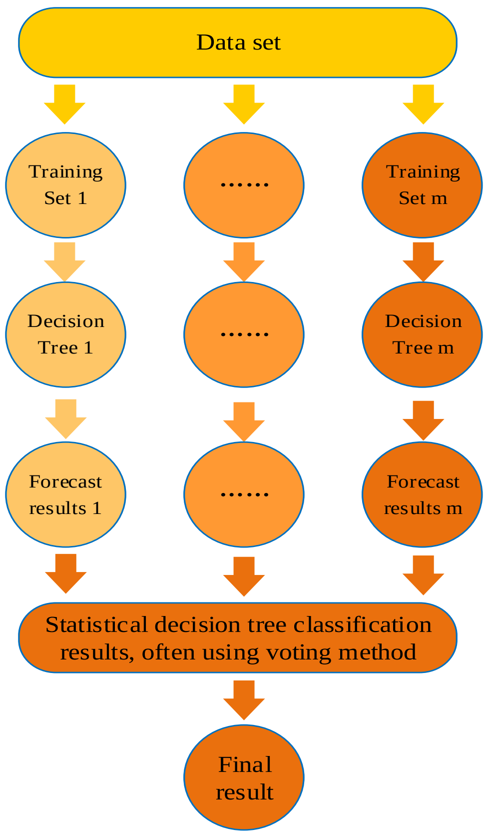


Figure 2: RF Model with multiple learners

Figure 2 shows that it first establishes multiple DT, each of which independently classifies the data. Finally, the output of these DT is collected, and the category with the highest score is designated. It is the final classification result. In the RF algorithm, the generation rules of the DT are divided into three steps: determining the training set of the DT, determining features. Finally, the DT is not pruned. If there are N samples in the training set, each DT needs to extract samples N times and put them back. This ensures that each training set has a partial intersection and is different. The classification ability of each DT and the correlation between DT determine the classification effectiveness of RF. In the RF algorithm, each DT is not pruned, which can make the RF less prone to overfitting problems and make the model have good noise resistance.

Adaboost algorithm is an important ensemble learning method in Boosting, which has the advantages of high accuracy and low overfitting. The algorithm is implemented through the following steps: first, setting the weights of the initial data to 1/N for N samples. It takes the error rate as an indicator and selects a weak classifier to train the data. After each training is completed, the sample weight of the sample set is updated and enters the next training. The classifier weights after each training session are calculated. Finally, the trained classifier is used to obtain the final classifier [17]. The process is shown in Figure 3.

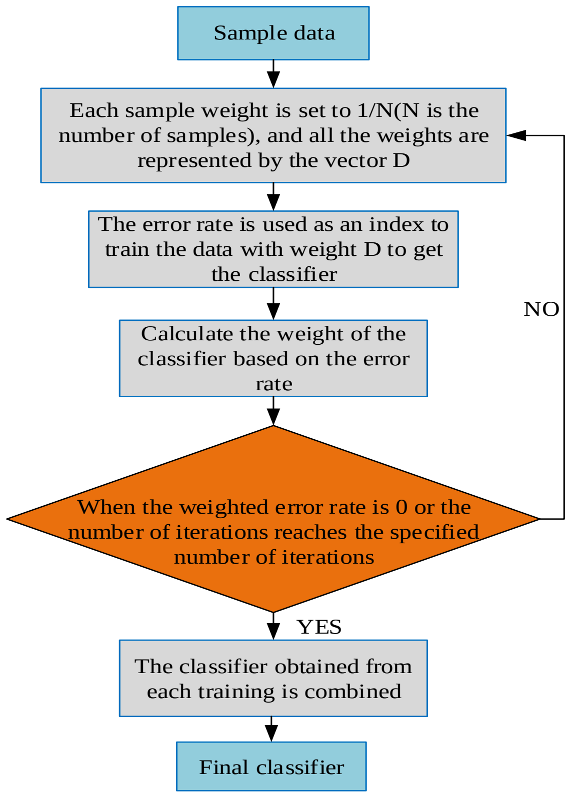


Figure 3: Adaboost algorithm process

LR is a commonly used binary classification method, implemented by linear regression and logical functions. The linear regression is showcased in equation (6).

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n \qquad (6)$$

In equation (6), x serves as the independent variable. θ serves as the coefficient. y represents the dependent variable. The relevant formula is showcased in equation (7).

$$g(y) = 1/(1+e^{-y}) \qquad (7)$$

In equation (7), y serves as the dependent variable. The process of implementing classification through LR is the process of transforming the results of linear regression into (0,1) using a logistic function. The predictive function of LR is shown in equation (8).

$$h_\theta(x) = g(\theta^T x) = 1/(1+e^{-}\theta^T x) \qquad (8)$$

In equation (8), x represents the independent variable. θ represents the coefficient. The prediction function of LR represents the probability of the result being classified as 1. When the input is x, the classification result is shown in equation (9).

$$\begin{cases} P(y=1|x;\theta) = h_\theta(x) \\ P(y=0|x;\theta) = 1-h_\theta(x) \end{cases} \qquad (9)$$

Equation (9) represents the probabilities when the classification result are 1 and 0, respectively, where $h_\theta(x)$ represents the prediction function of LR. The loss function is shown in equation (10).

$$\begin{cases} \mathrm{Cos}\,t(h_\theta(x), y) = -\log(h_\theta(x)), if\ \ y=1 \\ \mathrm{Cos}\,t(h_\theta(x), y) = -\log(1-h_\theta(x)), if\ \ y=0 \end{cases} \qquad (10)$$

Equation (10) represents the loss function value of the probability when the class result is divided into 1 and 0. The $J(\theta)$ is shown in equation (11).

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y(i)\log h_\theta(x^i) + (1-y^i)\log(1-h_\theta(x^i))] \qquad (11)$$

In equation (10), m represents the sample size. $h_\theta(x)$ represents the prediction function of LR. The minimum value of $J(\theta)$ is obtained through the gradient descent method, as shown in equation (12).

$$\theta_j := \theta_j - \alpha\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^i) - y^i)x^i{}_j, (j = 0,1,\ldots\ldots n) \qquad (12)$$

In equation (12), n represents the number of sample features. α represents the formula constant. Psychological crisis refers to the state in which an individual is unable to effectively cope with major life events or difficulties, resulting in serious psychological distress and pain [18]. This state may lead to abnormalities in individuals' emotions, cognition, behavior, and other aspects, which can have a negative impact on their MH and life. Psychological crisis may be caused by unexpected events, loss of important support systems, trauma, loss of family or friends, or significant changes. Common manifestations of psychological crisis include anxiety, depression, panic, suicidal tendencies, social disorders, etc.

3.2 Research on psychological crisis warning system based on fusion stacking model

This experiment selected students from a certain university and conducted a survey by distributing questionnaires. Data preprocessing is carried out on the data collected during the survey, to filter out data that is not suitable for the model from the original data, correct the data, or eliminate useless data. Data preprocessing is generally divided into three steps, as shown in Figure 4.

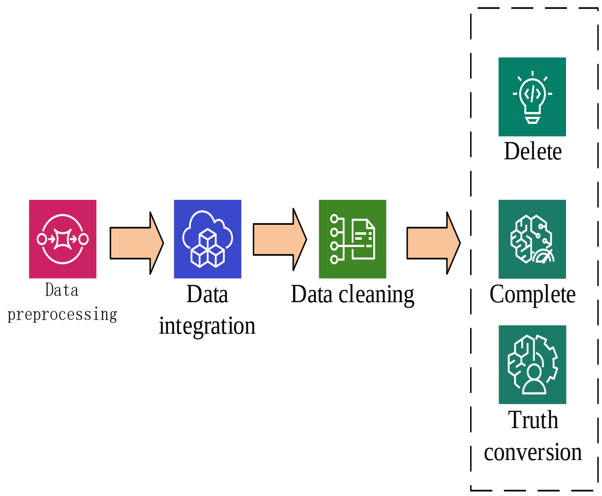


Figure 4: Data preprocessing steps

Figure 4 shows that data integration refers to the process of integrating data from multiple different sources for analysis and processing. In the process of data integration, it is necessary to address issues such as inconsistent data formats, data redundancy, and incomplete data. The data for this study mainly includes the results of personality tests for college students, which are analyzed by psychological workers on students who show abnormalities in the questionnaire.

This study uses standardized psychological scale thresholds and expert confirmation from certified school counselors to accurately label students' psychological states. Specifically, students are required to complete the Eysenck Personality Questionnaire (EPQ) and the Symptom Checklist-90 (SCL-90), both widely used in MH screening. A student is preliminarily marked as "psychologically abnormal". if their SCL-90 total score exceeds 160 points, or if any subscale (e.g., depression, anxiety, interpersonal sensitivity) scored ≥2.0 on average, the results are consistent with clinical screening thresholds. After initial screening, a panel of two university-affiliated psychological counselors independently review the questionnaire results and follow-up interviews for flagged cases. Only when both reviewers agree on the presence of a psychological crisis risk, the label "psychologically abnormal" is assigned. Cases without agreement are discarded from training. This two-step labeling process ensures high reliability of the label, although it is acknowledged that it reflects a quasi-clinical screening rather than a formal psychiatric diagnosis (i.e., not ICD-10/DSM-V certified). Therefore, the label represents a clinically-informed judgment based on validated scales and human oversight, serving as a proxy ground truth for supervised learning purposes.

In the questionnaire, there is a large amount of personal information. If not handled properly, it is highly likely to lead to the leakage of personal privacy. Based on the content of this study, it excluded information that was significantly unrelated to students' psychology, such as blood type, height, weight, etc. Then, one type is selected for storing duplicate information. Data cleaning refers to the processing and organization of collected raw data to eliminate errors, inconsistencies, and duplicates in the data, and achieve a reliable, consistent, and standardized state. The data in this study is mainly missing. The processing of missing values usually adopts four methods, namely deletion, completion, truth conversion, and no processing. Due to the fact that the psychological data used is mostly character data, it is necessary to convert character data into numerical data [19]. This study addresses this issue by encoding feature values in psychological archives. The RF algorithm and Adaboost model can achieve good results through label encoding. This is because the DT model is a process of finding the optimal splitting node, which requires calculating the information gain of the node. Therefore, the reduction and amplification of numerical values will not affect the splitting point, and the structure of the DT will not change. Therefore, this study uses label encoding to encode the data.

The evaluation of psychological abnormalities among college students is a binary classification problem, and common classification models include Bayesian model, DT model, support vector machine model, K-nearest neighbor model, LR model, etc. Before selecting a model, it is necessary to consider the characteristics of the data. This time, the psychological archive data has two characteristics: one is that the data has many related attributes; Another reason is that psychological abnormalities only account for a small portion of this data. Generally speaking, Bayesian models do not perform well in classification when there is a high correlation between attributes. DT models, RF models, and Adaboost models all perform well in non-uniform datasets.

When using DT algorithms for modeling, to prevent overfitting issues in the model, some parameters need to be set. The first is the depth of the DT. If the depth of the DT is too large, it will cause overfitting problems in the model. On the contrary, if the depth of the DT is too small, the model may have the problem of insufficient fitting. Therefore, it is necessary to set a maximum depth for the DT, and branches exceeding this maximum depth will be pruned. Secondly, the maximum number of features. If this parameter is set too small, it will lead to insufficient learning of the model. Finally, the minimum number of samples for leaf nodes. If this parameter is set too small, it can lead to overfitting problems. Setting it too small can result in the model not being able to learn the data well.

The RF model is an integrated algorithm based on DT model Bagging, which performs well in non-uniform datasets. When using RF for modeling, the most important parameter is the number of individual trees [20]. For RF models, the smaller the value, the easier the model is to underfit. The higher the value, the more excellent the model's performance, but the higher the value, the more excellent the computational power requirements of the model on the equipment.

In this study, a total of 200 college students from a Chinese university are selected as the experimental subjects. The data is collected using two components: A standardized questionnaire and a psychological game simulation. The questionnaire incorporates standardized psychological assessment scales, including the Eysenck Personality Questionnaire (EPQ) and Symptom Checklist-90 (SCL-90), covering dimensions such as neuroticism, extroversion, interpersonal sensitivity, and anxiety. The game simulation module is designed as a decision-based interactive scenario, where students are placed in virtual college life situations (e.g., peer conflict, academic pressure, financial difficulty, isolation) and are required to make choices under time constraints. The system records response latency, decision patterns, emotional keywords, and error rates. From these behavioral logs, 21 psychological and behavioral features are extracted, such as emotional reactivity score, impulsive decision frequency, choice consistency, and cognitive bias indicators. Then, all collected features are integrated into a unified student psychological profile and digitally encoded using tag coding techniques. This feature set is used as the input to train the RLA-Stacking model. The simulation component is based on Unity3D game engine and is embedded with rule-based stress-inducing mechanisms designed in collaboration with university psychology counselors. The game simulation data is included to address the shortcomings of self-report questionnaires by capturing implicit behavioral traits that are less prone to conscious manipulation.

Stacking is a common classification model fusion strategy that can perform multi-level fusion. Taking a two-layer approach as an example, the Stacking model fusion strategy involves using two models in one layer and then modeling and predicting them on initial training data. The output results of the model are integrated to form a new training dataset, and then the appropriate model is selected for modeling. Its structure is shown in Figure 5.

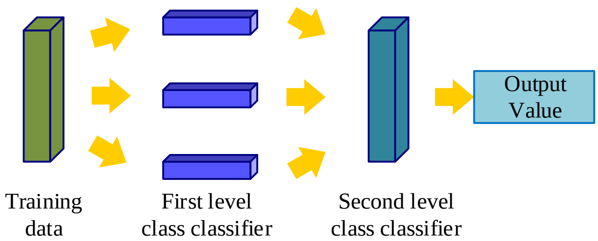


Figure 5: Structure of a two-layer Stacking fusion model

Multi-layer Stacking fusion is relatively rare because the performance of multi-layer fusion is not much better than that of two-layer fusion models, but its complexity and computational power consumption will greatly increase. Therefore, stacking fusion is generally two-layer [21]. Due to the need to select a base model for the Stacking fusion model, this study adopts the RF model, LR model, and Adaboost model as the base models. The fusion model used is the RLA-Stacking model, and its generation process is shown in Figure 6.

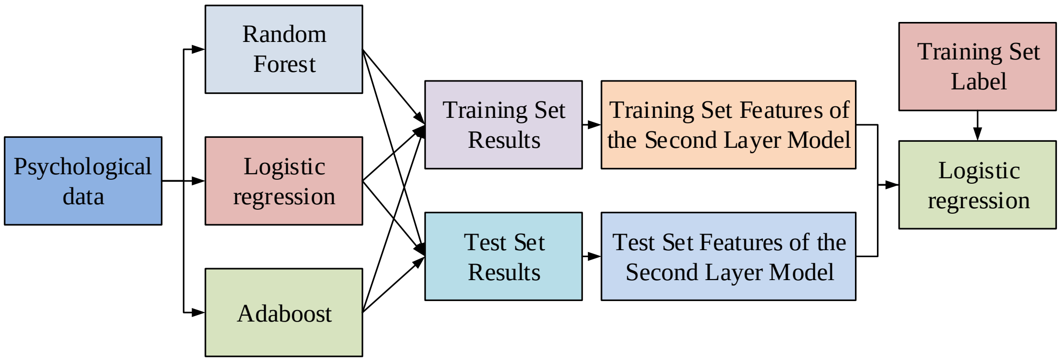


Figure 6: RLA-Stacking fusion model

Figure 6 shows that the psychological data is first separated into a training set and a testing set. The training set data is input into three models for training, and the prediction outcomes of the three models on the training set and testing set are output respectively. Using the predicted results as the training set features of the second layer model, the final result is obtained through LR [22].

The modeling process of the fusion prediction model based on the Stacking method is as follows. Firstly, it obtains the labels and feature sets of the data, and then analyzes the data to construct a psychological dataset. In the training of the test set, three models are trained using cross validation method, which can prevent overfitting problems. The main process is shown in Figure 7.

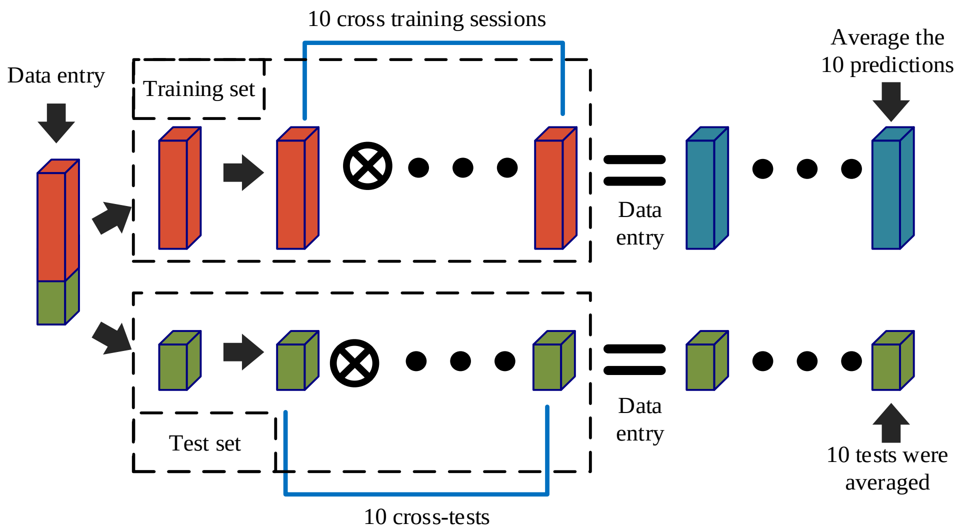


Figure 7: Cross-validation method

Figure 7 shows that after completing 10 cross-validations, the RF model will generate 20 prediction results, namely the training set and the testing set. It combines the prediction results on the training set to obtain the prediction results of the RF model. The results of the other two algorithms are outputted separately. The combination of the three results forms the feature set of the second layer model training set. It outputs the training set to the second layer model and trains the LR model. It evaluates the performance of the model by validating it through a validation set.

In the proposed RLA-Stacking model, three base classifiers are employed: RF, LR, and Adaboost. The RF model is configured with n_estimators = 100, max_depth = 10, and min_samples_split = 4. The Adaboost classifier uses a decision stump as its weak learner with n_estimators = 50 and learning_rate = 0.8. For LR, the L2 regularization is applied with a penalty term (C) of 1.0, and liblinear solver is used for small dataset optimization. All models are implemented using scikit-learn version 1.2.2. During the Stacking procedure, a 10-fold stratified cross-validation (StratifiedKFold) is applied on the training set to avoid data imbalance affecting model generalization. For each base learner, the training set is divided into 10-fold, 9-fold is used for training and the remaining one for generating predictions as meta-features, ensuring that no information leakage occurs between layers. This process is repeated for each fold, and the predictions are aggregated to form the full second-level training set. The same cross-validation folds are used consistently across the three base learners to maintain feature alignment. Finally, a LR classifier is trained as the second-level learner using these cross-validated predictions. During testing, each base learner is trained on the entire training set, and its prediction on the test set is used to construct the second-level test features. This ensures that the meta-model operates only on unseen data. This two-phase training and prediction strategy preserves model robustness and eliminates information overlap across model layers.

The term “game simulation” in this study refers to a behaviorally interactive simulation system rather than a gamified questionnaire or entertainment game. The system is developed using Unity3D and collaborated with licensed psychologists to design a series of scenario-based decision-making tasks. Each task mimics real-world stressors commonly encountered by college students, such as academic failure, peer exclusion, financial distress, or dorm conflicts. In each scenario, students must respond to a series of prompts within a limited time frame (e.g. 20-30 seconds for each selection) and record all actions taken. The system captures a variety of behavioral features including reaction time, impulsive decision patterns, choice switching, hesitancy, and emotional keyword selection. These features are later encoded as numerical indicators and combined with questionnaire results to form the full psychological profile used for model training. This simulation is neither a diagnosis nor a treatment, but rather serves as a non-invasive proxy for potential psychological characteristics, particularly useful for identifying risks without the need for explicit self-disclosure.

All procedures in this study are reviewed and approved by the Ethics Committee. The students are completely voluntary, and each participant has signed a written informed consent form before collecting data. The informed consent form clearly states that the data will only be used for scientific research, and personal identity will remain anonymous. Participants can withdraw from the study at any time without punishment. To ensure privacy, all psychological archives and behavioral logs collected from the game simulation and questionnaires are recorded under pseudonyms. Personally identifiable information such as student ID, name, or contact details is removed, and only encrypted numeric identifiers are retained. The data are stored on secure servers with access restricted to the research team. In addition, a referral protocol has been established: students marked as high-risk during the screening process are secretly contacted and provided with counseling resources at the university's MH center. These measures ensure compliance with ethical standards and protect the privacy and well-being of all participants.

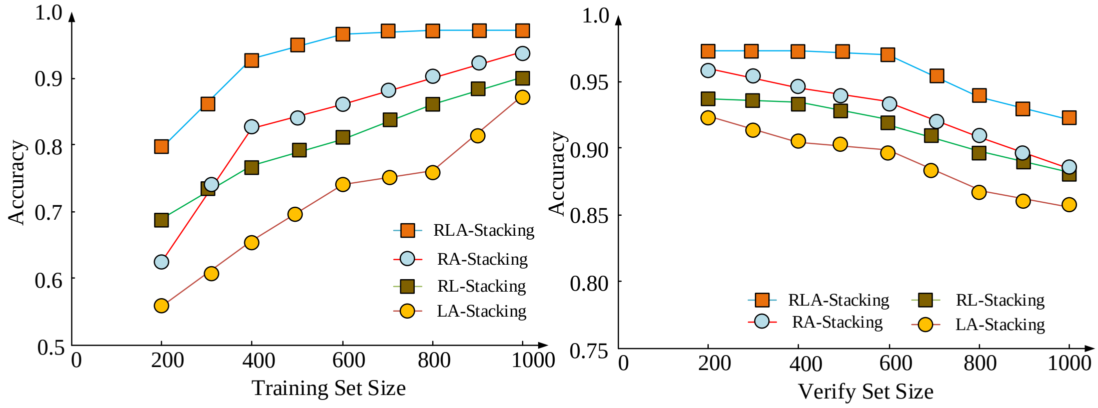
To better accommodate behavioral variability among students, adaptive mechanisms can be incorporated into the proposed framework. These adaptive features will enable the model to update its internal parameters in response to changing emotional or cognitive patterns observed during the game simulation. For example, the system can dynamically adjust feature weights, learning rates, or decision thresholds based on real-time deviations between predicted and observed behaviors. This adaptive adjustment will enable the model to capture the temporal drift of psychological states, enhance sensitivity to subtle behavioral changes, and maintain predictive accuracy under different academic periods or stress conditions. Therefore, integrating this adaptive learning strategy will make the psychological crisis warning system more flexible, personalized, and responsive to students' constantly changing psychological states.

4 Performance analysis of psychological crisis game simulation and warning system based on fusion stacking model

The first section of this chapter studied the performance of the model, introducing the LA-Stacking model, RL-Stacking model, and LA-Stacking model. The training and validation sets, iteration time, and other methods of the model were compared to verify the performance of the model; The second section analyzes the application functions of the model.

4.1 Performance analysis of psychological crisis warning system based on fusion stacking model

In the collected dataset, psychologically abnormal cases accounted for approximately 23% of the total samples, creating a notable class imbalance. To address this issue, several strategies were evaluated. Firstly, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to generate synthetic abnormal samples by interpolating between existing minority class instances. Secondly, random under-sampling was tested to reduce the number of majority class samples. Thirdly, class-weighting was enabled in all classifiers (RF, LR, and Adaboost), so that misclassifications for minority cases were penalized more heavily. Among these approaches, SMOTE combined with class-weighting provided the most stable results, improving the F1-score of the minority class by 6.4% and the AUC by 4.1% compared with the baseline without imbalance correction. These findings indicate that proper imbalance handling is essential to ensure that the proposed RLA-Stacking model maintains sensitivity to high-risk cases rather than being biased toward normal samples. The server CPU used in this study is Inter (R) Core (TM) i5-10210U, with 16GB of RAM, Windows 10 operating system, and 8GB of memory. The performance of different fusion models is compared on psychological data, as shown in Figure 8.

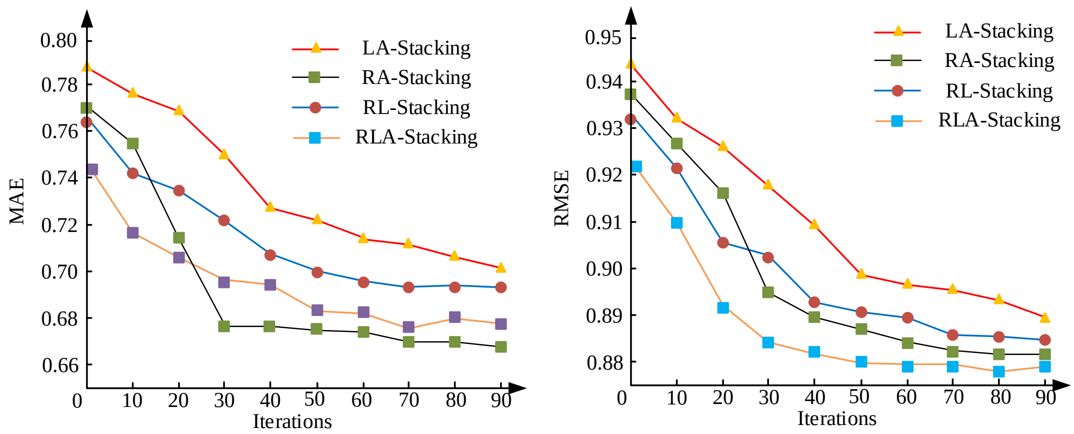


(a) Changes in accuracy of different algorithm models in training sets

(b) Changes in accuracy of different algorithm models in the validation set

Figure 8: Performance comparison of four hybrid models

Figure 8 (a) shows the changes in the accuracy of the four hybrid models as the training set grows. Figure 8 (b) shows the changes in the accuracy of the four hybrid models as the validation set increases. Figure 8 (a) shows that as the training set continues to increase, the accuracy of the LA-Stacking model, RL-Stacking model, LA-Stacking model, and RLA-Stacking model continues to increase. When the quantity of iterations is 400, the accuracy of the RLA-Stacking model basically reaches its maximum, while the accuracy of other models is still in the rising stage. When the training set size is 1,000, the accuracy of the RLA-Stacking model, LA-Stacking model, RL-Stacking model, and LA-Stacking model are 0.97, 0.95, 0.90, and 0.87. Figure 8 (b) shows that as the validation set increases, the accuracy of each model decreases, with the LA-Stacking model showing the greatest decrease in accuracy. When the validation set is 1,000, the accuracy of the RLA-Stacking model, LA-Stacking model, RL-Stacking model, and LA-Stacking model are 0.91, 0.88, 0.87, and 0.85. The experiment illustrates that the performance of the RLA-Stacking model can reach a high level when the training set is small, and its performance will not decrease significantly as the validation set is too large. To validate the performance of the model, the performance of the model is observed at different iterations, as shown in Figure 9.

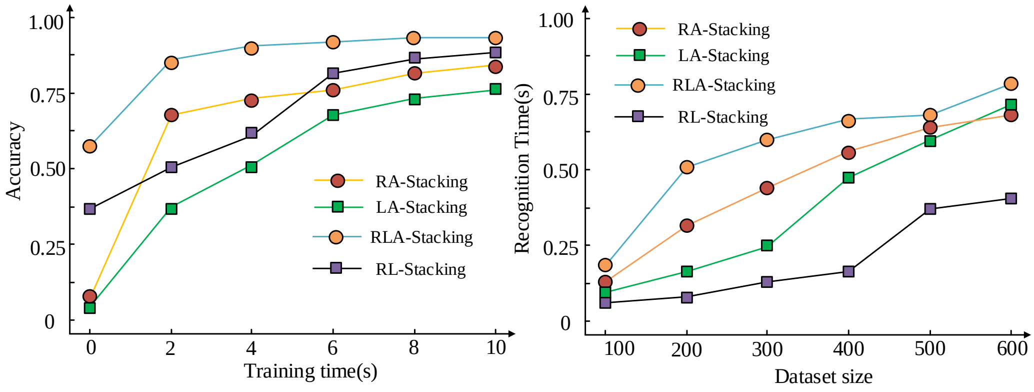


(a) MAE of Four Methods at Different Iterations

(b) RMSE of Four Methods at Different Iterations

Figure 9: Model performance observed in different iterations

Figure 9 (a) represents the MAE values of the four algorithms, while Figure 9 (b) represents the Root Mean Square Error (RMSE) values of the four algorithms. As shown in the figure, as the quantity of iterations increases, the MAE and RMSE values of the four models all decrease. When the number of iterations reaches 90, the MAE values of the LA-Stacking model, LA-Stacking model, RL-Stacking model, and RLA-Stacking model are 0.70, 0.69, 0.68, and 0.67, respectively. The RMSE values are 0.889, 0.885, 0.881, and 0.879. When the quantity of iterations of the RLA-Stacking model is around 30, the MAE and RMSE values of the model tend to stabilize, while other algorithms have significant fluctuations; The experiment showcases that the RLA-Stacking model can perform well at lower iterations, while other models require more iterations to perform well. This indicates that the performance of the proposed RLA-Stacking model is stronger than other models. It compares the response time and training time of the four models separately. The outcomes are showcased in Figure 10.



(a) The relationship between training time and accuracy (b) The relationship between recognition time and Dataset size

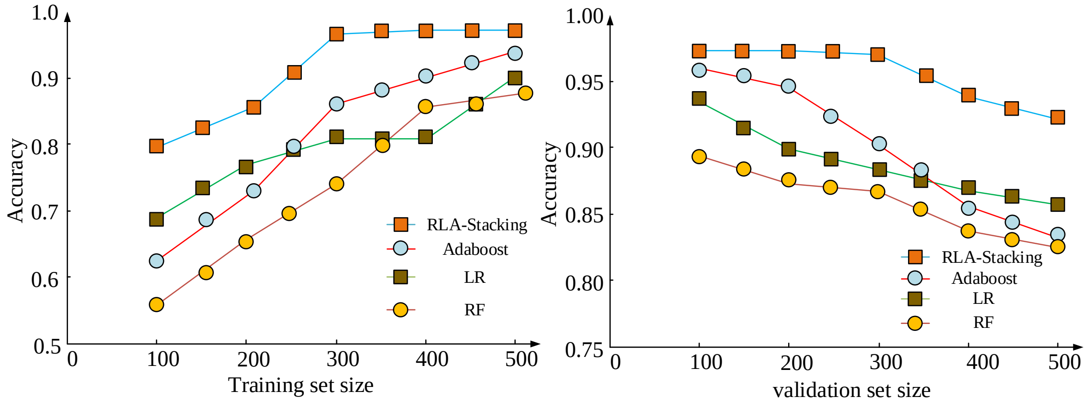
Figure 10: Response time and training time of four models

Figure 10 (a) shows the performance curves as the training time increases. Figure 10 (b) shows the response time of the four models as the dataset increases. As the training time grows, the accuracy of the model also grows. The accuracy of the RLA-Stacking model tends to be stable and can achieve high performance when the training time is about 2 seconds. The performance of other models tends to stabilize and reaches high performance only after the training time is about 6 seconds. Figure 10 (b) showcases that as the amount of data grows, the response time of the model also correspondingly grows. The RLA-Stacking model only increases its response time when the data size is 400, while other models begin to increase their response time when the data size is around 200. The experiment illustrates that the RLA-Stacking model has high accuracy in less training time, and its ability to process large amounts of data is significantly better than other models.

In addition to predictive accuracy, the computational efficiency was evaluated. The LR model completed training in approximately 2.1 seconds, Adaboost in 6.4 seconds, and RF in 9.8 seconds, while the proposed RLA-Stacking model required 15.3 seconds due to the two-layer training process. Despite the additional overhead, the runtime remained within practical limits for deployment. To further examine scalability, subsets of the dataset were generated by sampling 25%, 50%, 75%, and 100% of the data. As shown in Figure 10, training time increased nearly linearly with sample size across all models, indicating that the proposed ensemble method preserves computational tractability as the dataset grows. These results demonstrate that the RLA-Stacking model achieves improved accuracy with only moderate increases in computational cost.

4.2 Analysis of the application function of a psychological crisis warning system based on the fusion stacking model

To improve the generalization ability and stability of the prediction model in identifying students with psychological abnormalities, this study introduces three complementary basic learners, RF, LR, and Adaboost, for fusion construction. RF has strong anti noise and nonlinear modeling capabilities in random sampling of feature subspaces. LR has advantages in handling linear relationships and interpretability of feature weights, while Adaboost improves overall accuracy through iterative weak classifiers and is suitable for handling moderately imbalanced data. The combination of the three can enhance the modeling ability of complex feature interactions while retaining interpretability. In addition, a hierarchical k-fold cross-validation method is used to train and fuse the base learners to avoid overfitting caused by data bias. To validate the RLA-Stacking model, the RLA-Stacking model is compared with a single model, and the outcomes are shown in Figure 11.

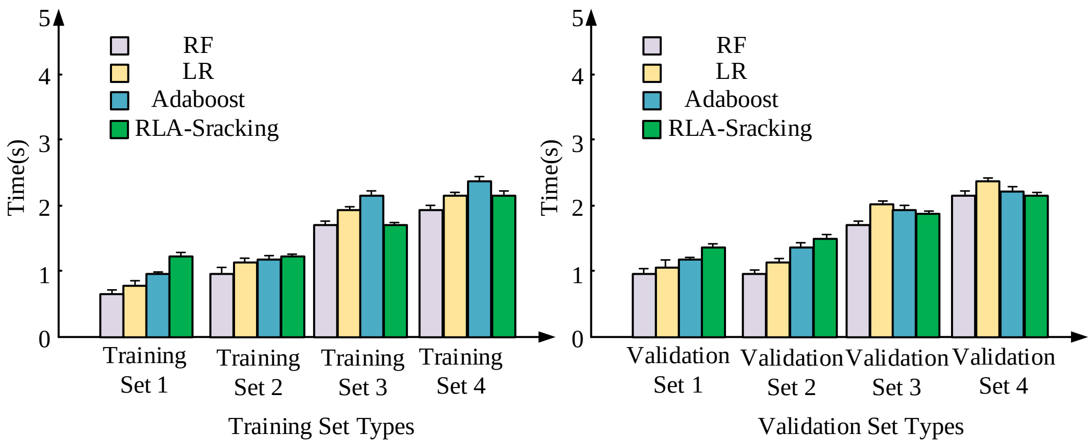


(a) Changes in accuracy of different models in The training set

(b) Changes in accuracy of different models in the validation set

Figure 11: Performance comparison between RLA-Stacking model and a single model

Figure 11 (a) shows the accuracy changes of RF, LR, Adaboost, and RLA-Stacking models as the training set grows. Figure 11 (b) shows the accuracy changes of RF, LR, Adaboost, and RLA-Tracking models as the validation set increases. As the training set increases, the performance also increases. When the training set size is around 300, the performance of the RLA-Tracking model tends to stabilize and exhibits the best performance. When the training set size is 500, the accuracy of RF, LR, Adaboost, and RLA-Tracking models is 0.88, 0.90, 0.95, and 0.98, respectively. As the validation set increases, the accuracy of all four models on the validation set decreases. The RLA-Tracking model only decreases when the validation set size is around 300, while the other three models begin to decrease when the validation set size is increased. The experiment showcases that the RLA-Tracking model can show better performance in a lower training set than other single models, and can show more stable performance in a larger data set. It selects training and validation sets of different sizes and divides them into 4 groups to compare the response time and training time of each algorithm. The results are shown in Figure 12.



(a)Training time for different models

(b)Response time for different models

Figure 12: Training time and response time of each model in different datasets

Figure 12 (a) shows the training time for different models on multiple datasets. Overall, the training time of the RLA-Stacking model showed a linear upward trend with the increase of data size, but its growth rate was significantly lower than that of traditional ensemble models such as Adaboost and RF. For example, on the largest dataset, the training time of RLA-Stacking was about 12.4 seconds, while RF reached 15.1 seconds and Adaboost was 16.3 seconds. Statistical tests ($F=9.52$, $p<0.01$) show significant differences between models, indicating that RLA-Stacking has better computational efficiency while maintaining high prediction accuracy. Figure 12 (b) reflects the response time of each model under the same data conditions. The results showed that as the amount of data increased, the response time of all models slightly increased, but RLA-Stacking remained at the lowest level. At the maximum sample size, its average response time was about 2.7 seconds, which was about 18% and 12% lower than that of Adaboost and LR, respectively. According to the analysis of variance

(F=8.64, $p<0.05$), the difference is statistically significant. The above algorithm models are compared, and the results are shown in Table 2.

Table 2: F1-scores of seven algorithms

Model type	Max F1 (%)	Min F1 (%)	Average F1 ± SD (%)
RLA-Stacking	83.1	80.1	81.8±0.9
RL-Stacking	83.4	77.5	80.7±1.1
LA-Stacking	80.7	74.6	78.4±1.3
RA-Stacking	81.6	78.9	79.6±1.0
RF	75.4	70.4	73.5±1.4
LR	80.8	74.6	78.3±1.2
Adaboost	81.9	75.7	80.1±1.0

Table 2 shows the F1-score of seven algorithms in psychological crisis prediction tasks, which can intuitively reflect the classification balance and stability of each model. From the overall results, the fusion model was superior to the single model. The average F1-score of RLA-Stacking algorithm reached 81.8 ± 0.9%, which was the best among all models, and the standard deviation was the smallest, indicating that it not only has high accuracy, but also has small fluctuations in the results, and has good robustness. The average F1-score of RL-Stacking and LA-Stacking was 80.7 ± 1.1% and 78.4 ± 1.3% respectively, which was slightly lower than that of RLA-Stacking, indicating that the performance is slightly reduced without the collaborative fusion of three models. The performance of traditional single models was relatively limited, with an average F1-score of only 73.5 ± 1.4% for RF, and a large standard deviation, indicating that the model is sensitive to data noise and sample imbalance. The average F1-scores of LR and Adaboost were 78.3 ± 1.2% and 80.1 ± 1.0%, respectively. Although slightly lower than the fusion model, they still maintain high stability. It is worth noting that the RA-Stacking model has a small difference between its maximum value of 81.6% and minimum value of 78.9%, with an average value of 79.6 ± 1.0%, indicating good consistency in different validation compromises. Overall, the fusion of multiple algorithms significantly improves the prediction accuracy and stability of the model, especially RLA-Stacking, which effectively balances bias and variance by integrating the complementary advantages of RF, LR, and Adaboost. It demonstrates better generalization ability and reliability than traditional single models in psychological crisis detection scenarios. It randomly selected 50 people from 4 schools and counted the number of patients. The number of patients was predicted using four models, and the outcomes are showcased in Table 3.

Table 3: Prediction of the number of patients with different models

School	RF (mean ± SD)	LR (mean ± SD)	Adaboost (mean ± SD)	RLA-Stacking (mean ± SD)	Actual Number
1	4.1±0.3	6.2±0.5	7.0±0.4	9.1±0.6	10
2	6.3±0.5	8.9±0.6	9.1±0.5	12.8±0.8	15
3	5.2±0.4	9.8±0.7	8.9±0.5	13.9±0.9	15
4	3.2±0.2	5.1±0.4	6.2±0.4	8.0±0.5	8

Table 3 compares the accuracy and stability of different models in predicting the number of psychologically abnormal students in four schools, reflecting the application effect of the model at the actual sample level. From the overall trend, the prediction results of the RLA-Stacking model are closest to the true values and have the smallest standard deviation, demonstrating the highest fitting accuracy and stability. For example, in School 1, RLA-Stacking predicted 9.1 ± 0.6 people, only 0.9 individuals different from the actual 10 individuals, while the predicted values of RF and LR were 4.1 ± 0.3 individuals and 6.2 ± 0.5 individuals, respectively, with significant errors. In School 2 and School 3, RLA-Stacking predicted 12.8 ± 0.8 and 13.9 ± 0.9 individuals, respectively, which were close to the actual 15 individuals, significantly better than that of other models. Although LR reached 9.8 ± 0.7 individuals in School 3, the trend of underestimation was still evident. The results of School 4 further demonstrated the advantages of the fusion model, with RLA-Stacking predicting 8.0 ± 0.5 individuals, which was completely consistent with the true value of 8 individuals, while Adaboost and RF underestimated by about 2 individuals. In terms of standard deviation, RLA-Stacking was less than 1 in all calibrations, indicating that its output stability is high in multiple repeated experiments and is not easily affected by sample distribution or random initialization. In contrast, RF has lower predicted values and the largest mean square error among all schools, indicating its limited performance in limited sample size and strong non-linear correlation features. The comprehensive results indicate that RLA-Stacking effectively improves the accuracy and consistency of predicting psychological crisis populations through multi-model integration, and can more reliably reflect the distribution of psychological risks among students in different schools, providing more valuable quantitative basis for subsequent warning interventions. It randomly selected 50 college students and divided them into 5 groups on average. The ratings of students on the model were collected, as shown in Figure 13.

Figure 13 shows that the RLA-Stacking model has high ratings among all groups of users, with five groups rating the RLA-Stacking model at 98, 92, 80, 91, and 78, Adaboost model at 86, 90, 71, 82, and 70, LR model at 89, 91, 75, 77, and 69, and Adaboost model at 84, 86, 68, 71, and 62, respectively. The experimental results show that each group of users has a higher evaluation of the RLA-Stacking model than other models, indicating that the proposed model has higher user preference.

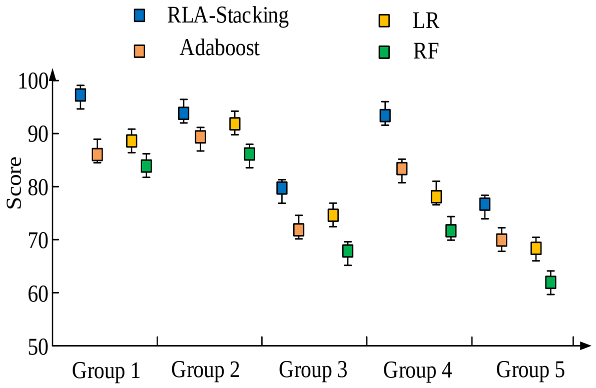


Figure 13: User rating chart

To enhance the practicality and long-term robustness of the proposed game-based psychological crisis warning system, several extensions are recommended. Firstly, feedback loops can be introduced to allow continuous parameter updates, enabling the model to learn from new behavioral data and refine its predictions over time. Secondly, neural-network-based enhancement modules may be incorporated to handle uncertainty and nonlinear interactions in multi-source behavioral features, thereby improving adaptability to diverse student groups. Thirdly, optimization techniques, such as nonlinear or meta-heuristic algorithms, can be applied to balance model accuracy, computational efficiency, and scalability for large institutional datasets. Finally, the framework should be validated with diverse real-world datasets from multiple universities to ensure its generalization and robustness under varying environmental and cultural conditions. Integrating these improvements will make the system more practical, flexible, and consistent with the most advanced adaptive control methods in intelligent behavior prediction.

The practical deployment of the proposed game-based psychological crisis warning system should align with university counseling frameworks. Behavioral data collected through interactive simulations can be securely stored in students' psychological profiles and reviewed by counselors under strict privacy protection. All participants must provide informed consent, and data anonymization or encryption should comply with institutional and legal standards. The model output should be interpretable, presenting key behavioral indicators that support professional evaluation, rather than replacing it. Although game simulation methods are innovative, they may introduce bias if students are detached from the real environment or behave differently from the real environment. This can be mitigated by repeated testing or combining behavioral data with self-reports. Compared to existing mobile MH tools and digital warning systems that primarily rely on questionnaires or social media data, the proposed framework provides a more interactive and ecologically effective pathway for continuous psychological monitoring, although large-scale deployment will require ethical supervision and interdisciplinary collaboration.

5 Discussion

In summary, the above experimental results demonstrate that the RLA-Stacking model possesses better prediction performance than single models such as RF, LR, and Adaboost. As shown in Table 2, the average F1-score of the RLA-Stacking model is 81.8%, which is higher than that of the LR model (78.3%), RF model (73.5%), and Adaboost model (80.1%). In Table 3, the predicted number of psychological crisis cases in four schools using the RLA-Stacking model is also closer to the actual values, indicating that this model has better generalization ability in real scenarios. Compared with the individual models, the performance improvement of the RLA-Stacking model mainly comes from the comprehensive advantages of the base classifiers. The RF model enhances robustness by constructing multiple DTs, the LR model ensures linear interpretability and simplicity, and the Adaboost model improves focus on misclassified samples. By combining their outputs through a second-level LR model, the RLA-Stacking model can fully integrate feature learning ability and classification stability, effectively reducing both bias and variance. In addition, compared to the existing research shown in Table 2 of the relevant work section, most models rely solely on psychological questionnaire data and do not consider the possibility of response bias or deception. In this study, a simulation-based game method is used to collect psychological behavior data, which reduces the chance of intentional concealment and enhances privacy protection. Therefore, from both the data source and model structure perspectives, the RLA-Stacking model shows better application potential and prediction effectiveness in psychological crisis identification.

While the current experiments confirm the stability and superior performance of the RLA-Stacking model, further validation is necessary to strengthen its generalizability. The model should be tested on external benchmark datasets or compared with state-of-the-art frameworks in psychological crisis detection, such as transformer-based or multi-modal deep learning systems that integrate linguistic, behavioral, and physiological features. Cross-validation on independent datasets from different institutions will help verify robustness across diverse student populations. In addition, longitudinal data integration can dynamically track psychological changes over time, providing more accurate trend prediction and early intervention cues. Future enhancements may incorporate behavioral analytics and real-time monitoring modules, allowing continuous data collection through digital interactions while maintaining ethical and privacy safeguards.

Compared with existing university-based psychological crisis prediction methods using social-media or sensor data, the proposed RLA-Stacking model achieves higher accuracy, robustness, and scalability [23-24]. It improves the average F1-score by 4-8% over single classifiers such as RF, LR, and Adaboost, and reduces F1-score by about 25 %, showing greater stability. The two-layer ensemble effectively captures linear and nonlinear feature dependencies, enhancing generalization and preventing overfitting. Runtime tests indicate near-linear training growth with increasing data, confirming computational efficiency for large-scale datasets. However, as it relies on interactive game simulations for behavioral data, deployment requires digital infrastructure, limiting use in resource-constrained environments. Overall, combining active behavioral elicitation with ensemble fusion offers an interpretable, ethical, and scalable framework for psychological crisis early warning in higher education.

6 Conclusion

With the improvement of quality of life, more and more people are paying attention to MH issues, among which the MH issues of college students are particularly important in society. This study proposes a stacked fusion model based on RF, LR, and iterative algorithms. This model is based on RF, LR, and iterative algorithms. It collects student information through game simulations, establishes students' psychological files, and analyzes the files to determine whether students have a psychological crisis. The experimental results show that when the validation set is 1,000, the accuracy of the RLA-Stacking model, LA-Stacking model, RL-Stacking model, and LA-Stacking model are 0.91, 0.88, 0.87, and 0.85. When the number of iterations reaches 90, the MAE values of the LA-Stacking model, LA-Stacking model, RL-Stacking model, and RLA-Stacking model are 0.70, 0.69, 0.68, and 0.67, respectively. The RMSE values are 0.889, 0.885, 0.881, and 0.879, respectively. The maximum F1-scores for RLA-Stacking, RL-Stacking, LA-Stacking, LA-Stacking, RF, LR, and Adaboost models are 83.1%, 83.4%, 80.7%, 81.6%, 75.4%, 80.8%, and 81.9%, respectively. The minimum F1-scores are 80.1%, 77.5%, 74.6%, 78.9%, 70.4%, 74.6%, and 75.7%, respectively. The average F1-scores are 81.8%, 80.7%, 78.4%, 79.6%, 73.5%, 78.3%, and 80.1%. This study randomly selects 50 college students to evaluate the model. The five groups rate the RLA-Stacking model at 98, 92, 80, 91, and 78, respectively. The Adaboost model is rated at 86, 90, 71, 82, and 70, while the LR model is rated at 89, 91, 75, 77, and 69. The Adaboost model is rated at 84, 86, 68, 71, and 62. The outcomes demonstrate that the proposed model possesses good model performance in predicting psychological crisis among college students.

The limitation of this study is that the dataset is collected from a single university, which may limit the generalizability of the research results to broader student populations. The differences in cultural backgrounds, institutional environments, and student demographic data among different universities may affect the psychological characteristics and crisis patterns captured by the model. Therefore, although the proposed RLA-Stacking model exhibits strong predictive performance in this situation, caution should be exercised when extrapolating the results to other institutions. Future research will focus on validating the model using multi-institutional and cross regional datasets, which will comprehensively evaluate the model's robustness and adaptability. This verification will also support the development of a scalable and widely applicable psychological crisis warning system.

The proposed game-based psychological crisis warning system extends beyond conventional questionnaire-driven assessments by introducing a data-driven, interactive, and ensemble-learning framework. To further improve adaptability and stability, future extensions can draw inspiration from advanced control theories. Specifically, adaptive fuzzy and backstepping control concepts can be employed to dynamically adjust prediction parameters as students' behavioral patterns evolve, enabling the model to maintain accuracy under non-stationary conditions. Incorporating output-feedback mechanisms may introduce closed-loop regulation, allowing the system to continuously refine its predictions based on observed deviations, analogous to stability enhancement in nonlinear dynamic control. In addition, robust neural adaptive control strategies can be integrated to manage uncertainty in multi-source behavioral data, and nonlinear optimal control techniques can serve as meta optimization tools to balance computational costs and performance. Integrating these adaptive and feedback-oriented control principles will endow the RLA-Stacking framework with greater flexibility, real-time responsiveness, and scalability, thereby promoting its practical application in complex educational environments.

References

[1] Gu T, Li J, Wang M, Duan P. Landslide susceptibility assessment in Zhenxiong County of China based on geographically weighted logistic regression model. Geocarto Int, vol. 37, no. 17, pp. 4952-4973. 2022. DOI:10.1080/10106049.2021.1903571.

[2] Gelete G, Gichamo T, Abraham T. Hybrid emotional neural networks and novel multi-model stacking algorithms for multi-lake water level fluctuation modeling. Earth Sci Inform, vol. 18, no. 2, pp. 121-135. 2025. DOI:10.1007/s12145-025-01733-z.

[3] Kolenik T, Schiepek G, Gams M. Computational psychotherapy system for mental health prediction and behavior change with a conversational agent. Neuropsychiatr. Dis. Treat., vol. 20, pp. 2465-2498, 2024. DOI: 10.2147/NDT.S417695

[4] Kolenik T. Intelligent cognitive system for computational psychotherapy with a conversational agent for attitude and behavior change in stress, anxiety, and depression. Informatica, vol. 49, no. 2, pp. 451-454, 2025. DOI: 10.31449/inf.v49i2.8738

[5] Li J. Analysis of the Mental Health of Urban Migrant Children Based on Cloud Computing and Data Mining Algorithm Models. Sci. Program, vol. 21, no. 7, pp. 763-785, 2021. DOI: 10.1155/2021/7615227

[6] Zishuo Zhu, Chang Su. From Games to Education: Research on Immersive Experience-Based Interactive Design of Children'S Educational Games. Advances In Industrial Engineering and Management, 2023, 12(1).

[7] Marques G, Drissi N, Isabel de la Torre Díez, Abajo B, Ouhbi S. Impact of COVID-19 on the psychological health of university students in Spain and their attitudes toward Mobile mental health solutions. Int. J. Med. Inf, vol. 147, no. 5, pp. 1043-1058, 2021. DOI: 10.1016/j.ijmedinf.2020.104369

[8] Sun X, Z.-J. W, Y.-Y. L, Chan K, Miao X, Zhao S, Wu Y, Li Z, Wu B. Corrigendum to "Trends of college students' mental health from 2005 to 2019 and its rural–urban disparities in China". Journal of affective, vol. 45, no. 306, pp. 354-361, 2021. DOI: 10.1016/j.jad.2022.03.048

[9] Chu Y, Yin X. Data Analysis of College Students' Mental Health Based on Clustering Analysis Algorithm. Complexity, vol. 12, no. 3, pp. 954-965, 2021. DOI:10.1155/2021/9996146.

[10] Liang F, Li P, Peng H, Ding B. Evaluation Model of College Students€ Mental Health Quality Based on Computational Intelligence. Math. Probl. Eng, vol. 2022, no. 7, pp. 1-11, 2022. DOI:10.1155/2022/1646082

[11] Jianhuan T Y S. Prediction and analysis of College Students' mental health based on BP neural network. J INTELL FUZZY SYST, vol. 47, no. 6, pp. 1-7, 2022. DOI:10.3233/JIFS-189359.

[12] Shao G, Han W, Zhang H, Liu S, Wang Y, Zhang L, Cui X. Mapping maize crop coefficient Kc using random forest algorithm based on leaf area index and UAV-based multispectral vegetation indices. Agric. Water Manage, vol. 25, no. 7, pp. 546-568, 2021. DOI:10.1016/j.agwat.2021.106906.

[13] Zhang Z, Cai Z. Permeability Prediction of Carbonate Rocks Based on Digital Image Analysis and Rock Typing Using Random Forest Algorithm. Sustainable Energy Fuels, vol. 35, no. 14, pp. 11271-11284, 2021. DOI: 10.1021/acs.energyfuels.1c01331.

[14] Manivel T, Saravanakumar U. Automatic digital analysis system to grade diabetic retinopathy by integrated stacking model concept. Traitement du Signal, vol. 41, no. 6, pp. 3275-3277, 2024. DOI:10.18280/ts.410643

[15] Li C, Xu B, Chen Z. A stacking model-based classification algorithm is used to predict social phobia. Appl Sci-Basel, vol. 14, no. 1, pp. 14-27, 2024. DOI:10.3390/app14010433.

[16] Bhagat N K, Mishra A K, Singh R K, Sawmliana C, Singh P.Application of logistic regression, CART and random forest techniques in prediction of blast-induced slope failure during reconstruction of railway rock-cut slopes. Eng. Fail. Anal, vol. 137, no. 35, pp. 1974-1987, 2022. DOI: 10.1016/j.engfailanal.2022.106230.

[17] Liu L, Bai F, Su C, Ma C, Yan R, Li H, Sun Q. Forecasting the occurrence of extreme electricity prices using a multivariate logistic regression model. Energy, vol. 247, no. 3, pp. 1275-1287, 2022. DOI:10.1016/j.energy.2022.123417.

[18] John L H, Kors J A, Reps J M, Ryan P, Rijnbeek P. Logistic regression models for patient-level prediction based on massive observational data: Do we need all data? Int. J. Med. Inf, vol. 163, no. 6, pp. 1047-1057, 2022. DOI: 10.1016/j.ijmedinf.2022.104762.

[19] Ostrovski V, Datta S, Koul H L. Testing equivalence to binary generalized linear models with application to logistic regression. Stat. Probab. Lett, vol. 191, no. 9, pp. 1845-1852, 2022, 191. DOI: 10.1016/j.spl.2022.109658

[20] Li Z, Zhao N, Zhang H, Wei Y, Chen Y, Ma Y. Research on high spatiotemporal resolution of XCO2 in Sichuan Province based on stacking ensemble learning. Sustainability, vol. 17, no. 8, pp. 1-22, 2025. DOI:10.3390/su17083433

[21] Li Q, Wang X. Bayesian optimization of stacking ensemble learning model for HPC compressive strength prediction. Expert Syst Appl, vol. 288, no. 1, pp. 128281.1-128281.9, 2025. DOI:10.1016/j.eswa.2025.128281

[22] Chen Z. Research on internet security situation awareness prediction technology based on improved RBF neural network algorithm. JCCE, vol. 1, no. 3, pp.103-108, 2022. DOI: 10.47852/bonviewJCCE149145205514

[23] Boulkroune A, Zouari F, Boubellouta A. Adaptive fuzzy control for practical fixed-time synchronization of fractional-order chaotic systems. J. Vib. Control, 2025. DOI: 10.1177/10775463251320258.

[24] Boulkroune A, Hamel S, Zouari F, Boukabou A, Ibeas A. Output-feedback controller based projective lag-synchronization of uncertain chaotic systems in the presence of input nonlinearities. Math. Probl. Eng., vol. 2017, pp. 1-12, 2017. DOI: 10.1155/2017/8045803.